Zilversmit, M., P. O'Grady, M. Russello, R. DeSalle. 2001. High throughput sequencing protocols for a survey of genomic characters in the Family Drosophilidae. *Dros. Inf. Serv.* 84: 199-201.

High throughput sequencing protocols for a survey of genomic characters in the Family Drosophilidae.

Zilversmit, M., P. O'Grady, M. Russello, R. DeSalle. Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA.

Introduction

It is easy to argue that that any scientist working on the molecular biology of his or her organisms can benefit from knowledge of the full genome sequence. The methods required to generate a complete sequence, however, are prohibitive for an academic laboratory due to the expense, labor and time required. A solution to this problem is to do a survey of molecular characters throughout the genome, using either whole genes or large fragments sampled from all functional and linkage groups. The completed draft sequence of *D. melanogaster* (Adams *et al.*, 2000) has enabled us to begin this type of genome survey for a number of Drosophilid taxa in our lab. Below is a summary of the methods developed to generate this large volume of data efficiently and economically.

Methods

The high throughput system active in our lab now functions in two stages, primer design and sequence generation.

Primer Design

Oligonucleotide primers for *D. melanogaster* genes are designed from each chromosomal segment via GADFLY (Genome Annotation Database of *D. melanogaster*). These sequences, and those from two or more homologous genes from other species (links found through GADFLY), are then input to the CODEHOP web site. CODEHOP produces ortholog blocks of sequence and a series of potential primers using the genetic code and codon bias tables. Optimal primers sequences are selected based on the following criteria:

(1) The primers should amplify a fragment between 300-800 bp. Shorter fragments are better if template DNA has been fragmented (due to sub-optimal storage of either the isolated DNA or the actual specimen), while longer targets are more economical, as 800 bp is the current maximum fragment size that can be sequenced with a single primer in both directions.
(2) Target sequences vary as to the level of sequence conservation between taxa, representing a range of conserved and highly variable regions.
(3) Primers are chosen with low to intermediate degeneracy (less than or equal to 32-fold).

In order to automate the sequencing step, we have incorporated the sequences for the T3 and T7 universal sequencing primers into our oligonucleotides (see Bonacum *et al*., 2001, this issue). This can be done either before or after they are tested for efficacy, though it is less expensive in the long run to redesign only the best primers. These universal primer sites are also located on the vector (TOPO pCR4; Invitrogen) used for cloning PCR products should that be necessary. With this addition only this

single set of primers is needed to sequence any number of different genes, either directly or ligated into a vector, and enables significantly greater throughput with little extra cost.

Primers are initially screened for success in PCR, assessing both for effectiveness in amplifying gene fragments from divergent taxa (*e.g.*, multiple diptera samples) and the number of products produced. Primers that produce a single product with the majority of test taxa, usually about 20% of tested primers, are selected to go on to the second stage where there are grouped by annealing temperature for use in PCR.

Sequence Generation

Using primers designed in the previous step, sequence data is generated by cycle sequencing with BigDye Terminator (ABI Prism) chemistry using PCR products from genomic DNA as template. It should be noted here that our methods are standardized for use with a an ABI Prism 3700 DNA Analyzer to generate sequence data; however, they can also be used with a lower throughput capillary or a slab gel sequencing system with little alteration. The only equipment requirements for the techniques are a repeater pipette, three multichannel pipettors (0.5-10µl, 5-50µl and 50-300µl volume capacity), a thermal cycler with a 96-well head, and a tabletop centrifuge with a plate rotor.

Template Preparation

Sequencing reaction templates are generated using a standard 30µl PCR with genomic DNA (isolated using a DNEasy Kit (Qiagen)) in a 96-well plate format.. The PCR products are purified by mixing them with SOPE Resin from the Quickstep 96 Well PCR Purification kit (Edge BioSystems) and then passing them through lab-made resin columns in the following manner. The Quickstep kit comes with a polystyrene cassette filled with resin columns, and though the resin can only be used once the 96-column cassette can be reused. The old columns are allowed to dry and are shaken out and the cassette is cleaned using filtered water. The column cells are then filled with 750µl of hydrated (1g to 14mL dH_20) Sephadex G50 beads (Pharmacia Biotech) each using a multichannel pipettor. To form the Sephadex columns the cassette is placed on a flat-bottom-well catch plate and spun twice on a tabletop centrifuge with a plate rotor at 2900rpm for 4 and 3 minutes respectively (with excess flow-through dumped between spins). At this point the columns are ready and 10µl of SOPE Resin is added to each PCR using a standard repeater pipette fitted with a 0.05mL Combitip (Eppendorf) (note: standard pipette tips cannot be used with this form of SOPE Resin, the Combitip can be rinsed with filtered water and reused for the same purpose). The mixture of PCR products plus SOPE Resin is then added to each column using a multichannel pipettor. The cassette is taped to a standard 96-well reaction plate and spun again at 2900rpm for 4 minutes. This system cleans PCR products for approximately 11.5 cents per reaction. The cleaned PCR products are then dried in a vacuum and resuspended with 20µl Rnase-free water each.

Sequencing Reactions

Sequencing reactions are prepared using BigDye Terminator (ABI Prism) diluted with equal parts Tris buffer (400mM Tris-HCl, pH9.0; 10mM $MgCl_2$). Purified PCR product template is added first, using a multichannel pipettor, and then the chemistry cocktail (primer, buffer, BigDye) using a repeater pipette (which is fitted with a 10µl attachment and using standard pipette tips). A reliable sequencing reaction can be completed using 2µl BigDye/buffer mix plus 2µl template DNA and 1µl 10µM universal primer. Sequences have been generated, however, using 1µl BigDye/buffer mix, or less. Completed sequencing reactions are cleaned using a standard isopropanol precipitation in the same

plate as the sequencing reactions. The reactions are then resuspended in 6µl of ABI deionized formamide and placed on an ABI 3700 for sequence generation.

Summary

1. *Primer Design*
   > Use GADFLY to walk along the *D. melanogaster* chromosomes
   > Use CODEHOP (Rose *et al*., 1998) to compare sequences
   (from the *C. elegans*, *D. melanogaster*, and human genomes when possible).
2. *Test Amplification*
   > Test primers with multiple taxa.
   > Test under a variety of amplification conditions.
3. *Redesign Primers with Universal Primers Sites*
   > Add universal primer sequence (T3 to F / T7 to R) to 5' end of degenerate primer
4. *Target Gene Amplification*
   > Amplify using PCR in 96-well plate
5. *PCR Purification*
   > Purify products in 96-well plates using SOPE Resin and lab-made columns
6. *Sequencing Reactions*
   > Standard protocols, cycle sequencing using 1/8 BigDye volume
7. *Sequencing Reaction Purification*
   > Purify in 96-well plates using alcohol precipitation
8. *Sequence Data Generation*
   > Resuspend samples in formamide, load on ABI 3700 DNA analyzer

References; Adams, M.D., *et al*., 2000, The genome sequence of *D. melanogaster*. *Science* 287: 2185-2215; Bonacum, J., *et al*., 2001, New nuclear and mitochondrial primers for systematics and comparative genomics in Drosophilidae. Dros. Inf. Serv. 84: this issue; Rose, T.M., *et al.*, 1998, Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 26:1628-1635.